Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models

Year: 2022 | Citations: 191 | Authors: Joseph L. Watson, David Juergens, N. Bennett, Brian L. Trippe, Jason Yim

Abstract

There has been considerable recent progress in designing new proteins using deep learning methods1–9. Despite this progress, a general deep learning framework for protein design that enables solution of a wide range of design challenges, including de novo binder design and design of higher order symmetric architectures, has yet to be described. Diffusion models10,11 have had considerable success in image and language generative modeling but limited success when applied to protein modeling, likely due to the complexity of protein backbone geometry and sequence-structure relationships. Here we show that by fine tuning the RoseTTAFold structure prediction network on protein structure denoising tasks, we obtain a generative model of protein backbones that achieves outstanding performance on unconditional and topology-constrained protein monomer design, protein binder design, symmetric oligomer design, enzyme active site scaffolding, and symmetric motif scaffolding for therapeutic and metal-binding protein design. We demonstrate the power and generality of the method, called RoseTTAFold Diffusion (RFdiffusion), by experimentally characterizing the structures and functions of hundreds of new designs. In a manner analogous to networks which produce images from user-specified inputs, RFdiffusion enables the design of diverse, complex, functional proteins from simple molecular specifications.